

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日

Date of Application: 2002年 7月15日

出願番号

Application Number: 特願2002-206202

[ST.10/C]:

[JP2002-206202]

出願人

Applicant(s): インターナショナル・ビジネス・マシーンズ・コーポレーション

2003年 2月 7日

特許庁長官  
Commissioner,  
Japan Patent Office

太田信一郎



出証番号 出証特2003-3005349

【書類名】 特許願

【整理番号】 JP9020093

【提出日】 平成14年 7月15日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 7/00

【発明者】

【住所又は居所】 神奈川県大和市下鶴間1623番地14 日本アイ・ビー  
ー・エム株式会社 東京基礎研究所内

【氏名】 安部 麻里

【発明者】

【住所又は居所】 神奈川県大和市下鶴間1623番地14 日本アイ・ビー  
ー・エム株式会社 東京基礎研究所内

【氏名】 小柳 光生

【発明者】

【住所又は居所】 神奈川県大和市下鶴間1623番地14 日本アイ・ビー  
ー・エム株式会社 東京基礎研究所内

【氏名】 小野 康一

【発明者】

【住所又は居所】 神奈川県大和市下鶴間1623番地14 日本アイ・ビー  
ー・エム株式会社 東京基礎研究所内

【氏名】 堀 雅洋

【発明者】

【住所又は居所】 神奈川県大和市下鶴間1623番地14 日本アイ・ビー  
ー・エム株式会社 東京基礎研究所内

【氏名】 仲池 卓也

【特許出願人】

【識別番号】 390009531

【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレ  
ーション

【代理人】

【識別番号】 100086243

【弁理士】

【氏名又は名称】 坂口 博

【代理人】

【識別番号】 100091568

【弁理士】

【氏名又は名称】 市位 嘉宏

【代理人】

【識別番号】 100108501

【弁理士】

【氏名又は名称】 上野 剛史

【複代理人】

【識別番号】 100104880

【弁理士】

【氏名又は名称】 古部 次郎

【手数料の表示】

【予納台帳番号】 081504

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9706050

【包括委任状番号】 9704733

【包括委任状番号】 0004480

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 データ処理方法、差分計算方法、これを用いた指示情報生成システム及びプログラム

【特許請求の範囲】

【請求項1】 コンピュータを制御して、構造化文書の所定の要素を指示するデータ処理方法であって、

所定の指示情報により要素を指示された構造化文書が変更された場合に、当該構造化文書を入力して変更内容を解析し、解析結果をメモリに格納するステップと、

前記メモリから前記解析結果を読み出し、解析された前記変更内容に応じて、変更後の前記構造化文書中の対応する要素を指示するように指示情報を更新するステップと

を含むことを特徴とするデータ処理方法。

【請求項2】 前記指示情報を更新するステップでは、XPathを用いて記述された前記指示情報を更新することを特徴とする請求項1に記載のデータ処理方法。

【請求項3】 コンピュータを制御して、2つのツリー構造データの差分を計算する差分計算方法であって、

メモリから処理対象である2つのツリー構造データを読み出して比較し、一方のツリー構造データから他方のツリー構造データに変形するための操作をツリー構造の構成要素に対する所定の操作の組合せで表現した操作列のリストを作成し、メモリに格納する第1のステップと、

前記メモリから前記操作列のリストを読み出し、当該操作列のリストのうちで、前記構成要素の移動として把握できる操作列を移動の操作に変更する第2のステップと

を含むことを特徴とする差分計算方法。

【請求項4】 前記第1のステップでは、前記ツリー構造データを変形するための操作を、ツリー構造のノードまたはサブツリーの追加、削除、変更の組合せで表現した操作列のリストを作成することを特徴とする請求項3に記載の差分

計算方法。

【請求項5】 複数の構造化文書の差分を計算する差分計算部と、

前記差分計算部により計算された差分に関する情報に基づいて所定の構造化文書における所定の部分を指示する指示情報から他の構造化文書中の対応する部分を指示する指示情報を生成する指示情報生成部と  
を備えることを特徴とする指示情報生成システム。

【請求項6】 構造化文書の構造を解析してツリー構造データに変換する文書解析部をさらに備え、

前記差分計算部は、前記文書解析部にて変換された複数の構造化文書に対応する複数のツリー構造データを比較して差分を計算することを特徴とする請求項5に記載の指示情報生成システム。

【請求項7】 前記差分計算部は、ツリー構造データの構成要素のうちで、前記複数のツリー構造データを一方から他方へ変形するための操作において移動された構成要素を追跡するように当該ツリー構造データの差分を計算することを特徴とする請求項6に記載の指示情報生成システム。

【請求項8】 前記指示情報は、X Pathで記述されていることを特徴とする請求項5に記載の指示情報生成システム。

【請求項9】 前記指示情報生成部は、前記構造化文書の差分と前記所定の構造化文書に対するX Pathとに基づいて、X Pathを構成するロケーションステップを再生成することにより、前記他の構造化文書に対するX Pathを生成することを特徴とする請求項8に記載の指示情報生成システム。

【請求項10】 コンピュータを制御して、構造化文書の所定の要素を指示するデータ処理を実行させるプログラムであって、

所定の指示情報により要素を指示された構造化文書が変更された場合に、当該構造化文書を入力して変更内容を解析し、解析結果をメモリに格納する第1の処理と、

前記メモリから前記解析結果を読み出し、解析された前記変更内容に応じて、変更後の前記構造化文書中の対応する要素を指示するように指示情報を更新する第2の処理と

を前記コンピュータに実行させることを特徴とするプログラム。

【請求項11】 前記プログラムによる前記第1の処理は、  
変更前後の前記構造化文書をツリー構造データに変換する処理と、  
当該ツリー構造データの差分を計算する処理とを含み、  
前記プログラムによる前記第2の処理では、前記ツリー構造データの差分を用  
いて前記指示情報を更新する処理を前記コンピュータに実行させることを特徴と  
する請求項10に記載のプログラム。

【請求項12】 前記プログラムによる前記差分を計算する処理では、前記  
構造化文書の変更に応じて変形されたツリー構造データの構成要素のうちで、当  
該ツリー構造データの変形に要する操作において移動された構成要素を追跡する  
ように当該ツリー構造データの差分を計算する処理を前記コンピュータに実行さ  
せることを特徴とする請求項11に記載のプログラム。

【請求項13】 前記プログラムによる前記第2の処理では、変更前後の前  
記構造化文書の差分に基づいて、前記指示情報が記述されたX Pathを構成す  
るロケーションステップを再生成することにより、当該X Pathの記述を更新  
する処理を前記コンピュータに実行させることを特徴とする請求項10に記載の  
プログラム。

【請求項14】 コンピュータを制御して、2つのツリー構造データの差分  
を計算するプログラムであって、

メモリから処理対象である2つのツリー構造データを読み出して比較し、一方  
のツリー構造データから他方のツリー構造データに変形するための操作をツリー  
構造の構成要素に対する所定の操作の組合せで表現した操作列のリストを作成し  
、メモリに格納する第1の処理と、

前記メモリから前記操作列のリストを読み出し、当該操作列のリストのうちで  
、前記構成要素の移動として把握できる操作列を移動の操作に変更する第2の処  
理と

を前記コンピュータに実行させることを特徴とするプログラム。

【請求項15】 前記プログラムによる前記第2の処理では、前記操作列の  
リスト中に、前記ツリー構造データにおける同一の構成要素を削除する操作と追

加する操作とが存在する場合に、これらの操作に替えて当該構成要素を移動する操作を当該操作列のリストに加える処理を前記コンピュータに実行させることを特徴とする請求項14に記載のプログラム。

【請求項16】 前記プログラムによる前記第2の処理では、前記操作列のリスト中に、前記ツリー構造データにおける所定の構成要素を変更する操作が存在する場合に、所定の規則に基づいて、当該操作を構成要素の移動を含む別の操作に置き替える処理を前記コンピュータに実行させることを特徴とする請求項14に記載のプログラム。

【請求項17】 コンピュータを制御して構造化文書の所定の要素を指示するデータ処理を実行させるプログラムを、当該コンピュータが読み取り可能に記録した記録媒体において、

前記プログラムは、

所定の指示情報により要素を指示された構造化文書が変更された場合に、当該構造化文書を入力して変更内容を解析し、解析結果をメモリに格納する処理と、

前記メモリから前記解析結果を読み出し、解析された前記変更内容に応じて、変更後の前記構造化文書における対応する要素を指示するように指示情報を更新する処理と

を前記コンピュータに実行させることを特徴とする記録媒体。

【請求項18】 コンピュータを制御して2つのツリー構造データの差分を計算するプログラムを、当該コンピュータが読み取り可能に記録した記録媒体において、

前記プログラムは、

メモリから処理対象である2つのツリー構造データを読み出して比較し、一方のツリー構造データから他方のツリー構造データに変形するための操作をツリー構造の構成要素に対する所定の操作の組合せで表現した操作列のリストを作成し、メモリに格納する処理と、

前記メモリから前記操作列のリストを読み出し、当該操作列のリストのうちで、前記構成要素の移動として把握できる操作列を移動の操作に変更する処理とを前記コンピュータに実行させることを特徴とする記録媒体。

【請求項19】 所定のH T M L 文書に対して設定されたアノテーションデータを管理するアノテーションサーバにおいて、

アノテーションデータが設定された前記H T M L 文書が変更された場合に、変更前後の当該H T M L 文書の差分を計算する差分計算手段と、

前記差分計算手段の計算により得られた差分情報に基づいて、前記アノテーションデータを前記H T M L 文書の所定の部分に対応付けるX P a t h を更新するX P a t h 更新手段と

を備えることを特徴とするアノテーションサーバ。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、X M L (Extensible Markup Language) やH T M L (Hypertext Markup Language) などで記述された文書の要素を指示する方法に関し、特に文書が変更された場合に要素の指示表現を更新する方法に関する。

【0002】

【従来の技術】

インターネットなどのデータ交換に用いられるX M L やH T M L などで記述された構造化された文書（以下、構造化文書）に対し、この構造化文書中の特定の要素（エレメント）を指示してアノテーション（annotation：注釈）などのメタ情報を付す場合がある。また、構造化文書に予め変形ルールを記述しておき、この変形ルールに基づいて構造化文書を変形させる場合がある。このようなメタ情報や変形ルールを構造化文書に付加する場合、X P a t h (XML Path Language) を用いて当該構造化文書における所定の位置を指示し、外部文書から参照することが行われる。

【0003】

X P a t h は、構造化文書の特定の部分を指示するための言語であり、X P a t h を指示情報として利用することによって、構造化文書中におけるアノテーションを付したり変形を行ったりする位置を任意に特定することができる。以下の説明では、X P a t h による記述も単にX P a t h と記す。

具体的な X P a t h の表記は、次のように行われる。

図18は、構造化文書であるXML文書の構成例を示す図である。X P a t h では、ルート要素をスラッシュ「/」で表す。したがって、図18のXML文書に対し、要素aは、ルートの子供要素であるため、「/a」と表現される。また、要素bや要素dは、「/a/b」や「/a/b/d」と表される。さらに、「//p[id="foo"]」は、あるXML文書中に存在する全てのp要素の中で、その要素のid属性が「foo」であるものを選択するX P a t h の表記例である。

#### 【0004】

##### 【発明が解決しようとする課題】

上述したように、X P a t h によれば、XML文書やH T M L文書といった構造化文書の特定の要素を任意に指示することができる。しかし、指示対象である構造化文書が変更されてしまうと、構造化文書中の要素やその位置が変わってしまうことにより、X P a t h による位置の指定が外れ、所望の要素を適切に指示できなくなる場合があった。

従来、このような場合に構造化文書中の所望の要素を適切に指示し続けるためには、人手を介してX P a t h の記述を修正しなければならず、多大な手間を要し、この構造化文書を利用するシステムの開発者にとって大きな負担となっていた。

#### 【0005】

そこで、本発明は、構造化文書中の所定の要素が指示されている場合に、この構造化文書が変更されても所望の要素を適切に指示し続けることができるようすることを目的とする。

また本発明は、X P a t h を用いて所定の要素が指示されている構造化文書が変更された場合に、この構造化文書の変更に応じてX P a t h を自動的に更新する手段を提供することを目的とする。

#### 【0006】

##### 【課題を解決するための手段】

上記の目的を達成する本発明は、構造化文書の所定の要素を指示する、次のようなデータ処理方法として実現される。すなわち、このデータ処理方法は、所定

の指示情報により要素を指示された構造化文書が変更された場合に、この構造化文書を入力して変更内容を解析するステップと、解析された構造化文書の変更内容に応じて、変更後の構造化文書中の対応する要素を指示するように指示情報を更新するステップとを含むことを特徴とする。

#### 【0007】

より詳細には、この構造化文書の変更内容を解析するステップは、変更前後の構造化文書をツリー構造データに変換する処理と、このツリー構造データの差分を計算する処理とを含む。そして、このツリー構造データの差分を用いて指示情報を更新する。

さらに詳しくは、このツリー構造データの差分を計算する処理では、この構造化文書の変更に応じて変形されたツリー構造データの構成要素のうちで、このツリー構造データの変形に要する操作において移動された構成要素を追跡するようこのツリー構造データの差分を計算する。

#### 【0008】

また好ましくは、構造化文書中の要素を指示する指示情報として X P a t h を用いることができる。

そして、指示情報を更新する処理では、変更前後の構造化文書の差分に基づいて、指示情報が記述された X P a t h を構成するロケーションステップを再生成することにより、 X P a t h の記述を更新する。

#### 【0009】

また、上記の目的を達成する他の本発明は、このようなデータ処理を実行する指示情報生成システムとしても実現される。すなわち、この指示情報生成システムは、複数の構造化文書の差分を計算する差分計算部と、この差分計算部により計算された差分に関する情報に基づいて所定の構造化文書における所定の部分を指示する指示情報から他の構造化文書中の対応する部分を指示する指示情報を生成する指示情報生成部とを備えることを特徴とする。

より好ましくは、この指示情報生成システムは、構造化文書の構造を解析してツリー構造データに変換する文書解析部をさらに備え、差分計算部は、この文書解析部にて変換された複数の構造化文書に対応する複数のツリー構造データを比

較して差分を計算する。

#### 【0010】

また、本発明は、このような指示情報の更新を実現するため、次のような2つのツリー構造データの差分計算方法として実現することができる。すなわち、この差分計算方法は、メモリから処理対象である2つのツリー構造データを読み出して比較し、一方のツリー構造データから他方のツリー構造データに変形するための操作をツリー構造の構成要素に対する所定の操作の組合せで表現した操作列のリストを作成するステップと、この操作列のリストのうちで、ツリー構造の構成要素の移動として把握できる操作列を移動の操作に変更するステップとを含むことを特徴とする。

ここで、ツリー構造の構成要素とは、ツリーのノード及び部分的なサブツリーを含む。また、ツリー構造の構成要素に対する所定の操作の組合せとは、例えばこれらの構成要素を追加し、削除し、または変更する操作の組合せといった、基本操作である。

#### 【0011】

また、より詳細には、このリスト中の操作列を変更するステップでは、このリスト中に、ツリー構造データにおける同一の構成要素を削除する操作と追加する操作とが存在する場合に、これらの操作に替えて構成要素を移動する操作をこのリストに加える。

さらに、このリスト中に、ツリー構造データにおける所定の構成要素を変更する操作が存在する場合に、所定の規則に基づいて、この操作を構成要素の移動を含む別の操作に置き替える。

#### 【0012】

また、上記の目的を達成するさらに他の本発明は、所定のH T M L文書に対して設定されたアノテーションデータを管理する、次のようなアノテーションサーバとしても実現される。このアノテーションサーバは、アノテーションデータが設定されたH T M L文書が変更された場合に、変更前後のかかるH T M L文書の差分を計算する差分計算手段と、この差分計算手段の計算により得られた差分情報に基づいて、アノテーションデータをH T M L文書の所定の部分に対応付ける

X Pathを更新するX Path更新手段とを備えることを特徴とする。

#### 【0013】

さらにまた、本発明は、コンピュータを制御して上記のデータ処理方法や差分計算方法における各ステップに対応する処理を実行させるプログラム、あるいは上述した指示情報更新システムやアノテーションサーバとしてコンピュータを機能させるプログラムとしても実現される。このプログラムは、磁気ディスクや光ディスク、半導体メモリ、その他の記録媒体に格納して配布したり、ネットワークを介して配信したりすることにより提供することができる。

#### 【0014】

##### 【発明の実施の形態】

以下、添付図面に示す実施の形態に基づいて、この発明を詳細に説明する。

図1は、本実施の形態によるX Path更新方法を実現するのに好適なコンピュータ装置のハードウェア構成の例を模式的に示した図である。

図1に示すコンピュータ装置は、演算手段であるCPU (Central Processing Unit: 中央処理装置) 101と、M/B (マザーボード) チップセット102及びCPUバスを介してCPU101に接続されたメインメモリ103と、同じくM/Bチップセット102及びAGP (Accelerated Graphics Port) を介してCPU101に接続されたビデオカード104と、PCI (Peripheral Component Interconnect) バスを介してM/Bチップセット102に接続されたハードディスク105、ネットワークインターフェイス106及びUSBポート107と、さらにこのPCIバスからブリッジ回路108及びISA (Industry Standard Architecture) バスなどの低速なバスを介してM/Bチップセット102に接続されたフロッピーディスクドライブ109及びキーボード/マウス110とを備える。

なお、図1は本実施の形態を実現するコンピュータ装置のハードウェア構成を例示するに過ぎず、本実施の形態を適用可能であれば、他の種々の構成を取ることができ。例えば、ビデオカード104を設ける代わりに、ビデオメモリのみを搭載し、CPU101にてイメージデータを処理する構成としても良いし、ATA (AT Attachment) などのインターフェイスを介してCD-ROM (Compact

Disc Read Only Memory) やDVD-ROM (Digital Versatile Disc Read Only Memory) のドライブを設けても良い。

#### 【0015】

図2は、図1に示したコンピュータ装置にて実現される本実施の形態によるX Path更新システムの構成を示す図である。

図2に示すように、本実施の形態によるX Path更新システムは、構造化文書の構造を解析する文書解析部10と、文書解析部10の解析結果に基づいて構造化文書に対してなされた変更を調べる差分計算部20と、差分計算部20の計算結果に基づいて指示情報としてのX Pathの記述を更新するX Path更新部30とを備えている。

#### 【0016】

上記の各構成要素は、図1に示したメインメモリ103に展開されたプログラムにてCPU101を制御することにより実現される仮想的なソフトウェアプロックである。CPU101を制御してこれらの機能を実現させる当該プログラムは、磁気ディスクや光ディスク、半導体メモリ、その他の記憶媒体に格納して配布したり、ネットワークを介して配信したりすることにより提供される。本実施の形態では、図1に示したネットワークインターフェイス106やフロッピーディスクドライブ109、図示しないCD-ROMドライブなどを介して当該プログラムを入力し、ハードディスク105に格納する。そして、ハードディスク105に格納されたプログラムをメインメモリ103に読み込んで展開し、CPU101にて実行することにより、図2に示した各構成要素の機能を実現する。

また、処理対象である構造化文書やX Pathは、例えばハードディスク105の所定の領域に格納されており、本実施の形態によるX Pathの更新処理を実行するために、CPU101によって読み出される。

#### 【0017】

図2に示した本実施の形態において、文書解析部10は、構造化文書を解析し、DOMツリーなどのツリー構造の形をしたデータ（以下、単にツリーと記す）に変換する。変換対象は、変更が行われた構造化文書における変更前後の各文書である。すなわち、文書解析部10は、ハードディスク105等の記憶手段に格

納されている処理対象の構造化文書（変更が行われた構造化文書）に関して、変更前の文書（以下、変更前文書）Pと変更後の文書（以下、変更後文書）P'を読み出して解析し、それぞれツリーT（変更前文書Pに対応）とツリーT'（変更後文書P'に対応）を生成し出力する。出力されたツリーT、T'は、メインメモリ103等の記憶手段に一時的に保持され、差分計算部20にて利用される。構造化文書のツリーを生成する変換アルゴリズムとしては、既存の従来技術を利用することができます。

#### 【0018】

差分計算部20は、文書解析部10により変換された、変更前後の各構造化文書（ツリー）の差分を計算する。これにより、処理対象である構造化文書に対してなされた変更の内容が認識される。本実施の形態では、差分の計算方法として、後に行うXPathの更新に適した新規な方法を提案する。以下、この差分計算アルゴリズムについて説明する。

まず、前提として、従来の一般的な差分計算のアルゴリズムを説明する。2つのツリーの差分を計算するアルゴリズムとしては、従来から種々のアルゴリズムが提案されているが、典型的な差分計算アルゴリズムに、最小コスト操作列を算出するものがある。

#### 【0019】

図16、17は、ツリーの差分計算の例を示す図である。

図16に示すように、ノードaを親とし、ノードb、c、dを子とするツリー構造で、ノードb、dの位置が入れ替わった2つのツリー161、162の差分を計算する場合を考える。なお、ツリーを変形する操作における処理コストは、基本操作である所定のノードを削除するRemoveNode、所定のノードを追加するInsertNode、ノードの内容を変更するModifyにおいて、それぞれ1とする。

このとき、最小コスト操作列を算出するアルゴリズムでは、ツリー161からツリー162へ変形するためには、ノードbの内容をノードdの内容に変更し（`Modify(b→d)`）、ノードdの内容をノードbの内容に変更する（`Modify(d→b)`）という操作が行われると計算する。これは、上記の処理コストの値から、処理コスト2でツリー161からツリー162へ変形することができ、これが最小の値

だからである。

### 【0020】

また、図17に示すように、ノードaを親とし、ノードb、c、dを子とするツリー構造で、ノードc、dに対するノードbの位置が異なる2つのツリー171、172の差分を計算する場合を考える。

このとき、最小コスト操作列を算出するアルゴリズムでは、ツリー171からツリー172へ変形するためには、ツリー171からノードbを削除し(DeleteNode(b))、ツリー172における図示の位置にノードbを追加する(InsertNode(b))という操作が行われると計算する。これも、上記の処理コストの値から、最小の処理コスト2でツリー171からツリー172へ変形することができるためである。

### 【0021】

これに対し、本実施の形態では、この差分計算がX Pathの自動更新のために行われることから、かかる最小コスト操作列を算出するアルゴリズムは好適な手法とは言えない。図16の例では、ツリー161に対してX Pathがノードbを指していたと仮定すると、ノードの内容が(bからdへ)変更されたという情報のみでは、X Pathの指示先を適切に変更することはできない。この例の場合、ツリー161からツリー162への変換は、ノードbとノードdとを入れ替えるように移動する操作が行われたと見ることができ、この情報に基づけば、ノードbを指すX Pathの記述を適切に変更することが可能である。したがって、X Pathを更新するための情報として用いるには、ツリー161からツリー162への変換を表す操作列は、ノードを移動する操作のMoveNodeを用いてMoveNode(b)、MoveNode(d)と表す方が適している。しかしながら、MoveNodeは、上述した基本操作DeleteNode、InsertNodeの組合せ(ノードを削除し、追加する操作)であるため、ツリー161からツリー162への変換に要するトータルコストは4となる。したがって、最小コスト操作列を算出するアルゴリズムではMoveNodeを検出することはできない。

### 【0022】

同様に、図17の例においても、ノードbを削除する操作またはノードbを追

加する操作の一方の情報から X P a t h の指示先を適切に変更することはできない。そして、ツリー 171 からツリー 172 への変換は、ノード b をノード c の左からノード d の右へ移動する操作が行われたと見ることができ、この情報に基づけば、ノード b を指す X P a t h の記述を適切に変更することが可能である。したがって、ツリー 171 からツリー 172 への変換を表す操作列は、MoveNode (b) と表す方が適している。この例では、MoveNode は、RemoveNode、InsertNode の組合せであることからトータルコストは 2 で変わらない。しかし、最小コスト操作列を算出するアルゴリズムで MoveNode を検出できる保証はなく、やはり本実施の形態には適さない。

なお、以上の考察は、サブツリー（ツリー中の部分的なツリー構造）を移動する操作の MoveTree についても同様である。以下では簡単のため、個々のノードに対する処理についてのみ説明するが、同様の解析を MoveTree に対しても行うことができるの言うまでもない。

### 【0023】

以上のこと踏まえて、差分計算部 20 にて実行される、本実施の形態に好適な差分計算アルゴリズムを説明する。本実施の形態で用いられる差分計算アルゴリズムは、ツリーの変形において移動されたもの（ノードやサブツリー）を追跡することを方針とする。

図 3 乃至図 6 は、本実施の形態に好適な差分計算アルゴリズムによる処理を説明するフローチャートである。

差分計算部 20 は、メインメモリ 103 等の記憶手段に一時的に保持されている、変更前文書 P に対応するツリー T と変更後文書 P' に対応するツリー T' とを入力し、基本操作である 3 つの操作 RemoveNode、InsertNode 及び Modify を用いて、ツリー T からツリー T' へ変形するために必要な操作を解析し、得られた操作列のリスト L を生成する。この解析には、上述した最小コスト操作列を算出するアルゴリズムなど、既存の手法を用いることができる。生成された操作列のリスト L は、メインメモリ 103 等の記憶手段に一時的に保持される。そして次に、差分計算部 20 は、メインメモリ 103 等に保持された操作列のリスト L に対して図 3 乃至図 6 に示す解析を行って、MoveNode を検出する。

## 【0024】

図3に示すInsertNode解析において、差分計算部20は、まずリストL中の所定のInsertNode (InsertNode(n))に着目する（ステップ301）。そして、着目したInsertNodeの対象ノード（図示の例ではノードn）と同一のノードに対する操作RemoveNodeがリストL中に存在するかどうかを調べる（ステップ302）。このとき、対応するRemoveNode（すなわち、RemoveNode(n)）が存在しないならば、ノードnは、ツリーT'において新たに追加されたノードであるので、そのまま処理を終了する。

一方、RemoveNode(n)が存在する場合は、これとInsertNode(n)との組合せによりMoveNode(n)が構成されるので、リストLにMoveNode(n)を追加し（ステップ303）、リストLからInsertNode(n)とRemoveNode(n)とを削除する（ステップ304）。差分計算部20は、以上の処理をリストL中の全てのInsertNodeに対して行う。

## 【0025】

図4に示すRemoveNode解析において、差分計算部20は、まずリストL中の所定のRemoveNode (RemoveNode(n))に着目する（ステップ401）。そして、着目したRemoveNodeの対象ノード（図示の例ではノードn）と同一のノードに対する操作InsertNodeがリストL中に存在するかどうかを調べる（ステップ402）。このとき、対応するInsertNode（すなわち、InsertNode(n)）が存在しないならば、ノードnは、ツリーT'において削除されたノードであるので、そのまま処理を終了する。

一方、InsertNode(n)が存在する場合は、これとRemoveNode(n)との組合せによりMoveNode(n)が構成されるので、リストLにMoveNode(n)を追加し（ステップ403）、リストLからRemoveNode(n)とInsertNode(n)とを削除する（ステップ404）。差分計算部20は、以上の処理をリストL中の全てのRemoveNodeに対して行う。

## 【0026】

図5及び図6に示すModify解析において、差分計算部20は、まずリストL中のノードn1の内容を変更する操作Modify(n1→nx)に着目する（ステップ501

)。そして、反対にノードの内容をn1にする変更を行う操作Modify(ny→n1)がリストL中に存在するかどうかを調べる(ステップ502)。

Modify(ny→n1)が存在するならば、次にノードnxの内容とノードnyの内容とが同一(すなわち、nx=ny)かどうかを調べる(ステップ503)。nx=nyである場合、ノードn1とノードnx(=ny)の位置を入れ替わったと解釈できるので、リストLにMovenode(n1)、Movenode(ny)を追加し(ステップ504)、リストLからModify(n1→nx)とModify(ny→n1)とを削除する(ステップ513)。

一方、nx=nyでない場合、ノードn1がツリーTにおけるノードnyのあった位置に移動し、当該ノードnyは削除され、元のノードn1の位置に他のノードnxが新たに追加されたと解釈できる。そこで、リストLにInsertNode(nx)、RemoveNode(ny)、Movenode(n1)を追加し(ステップ505)、さらに図3、4に示したInsertNode解析、RemoveNode解析を実行する(ステップ506)。この後、リストLからModify(n1→nx)とModify(ny→n1)とを削除する(ステップ513)。

#### 【0027】

また、ステップ502でModify(ny→n1)が存在しない場合、次に差分計算部20は、ノードn1に対する操作InsertNode(n1)がリストL中に存在するかどうかを調べる(図6、ステップ507)。存在するならば、ノードn1が他の所定の位置に移動し、元のノードn1の位置に他のノードnxが新たに追加されたと解釈できる。そこで、リストLにInsertNode(nx)、Movenode(n1)を追加し(ステップ508)、さらに図3に示したInsertNode解析、を実行する(ステップ509)。この後、リストLからModify(n1→nx)とInsertNode(n1)とを削除する(図5、ステップ513)。

#### 【0028】

さらに、ステップ507でInsertNode(n1)がリストL中に存在しない場合、次に差分計算部20は、ノードn1に対する操作RemoveNode(nx)がリストL中に存在するかどうかを調べる(図6、ステップ510)。存在するならば、ノードn1が削除され、同じ位置にノードnxが移動されたと解釈できる。そこで、リス

トLにRemoveNode(n1)、Movenode(nx)を追加し（ステップ511）、さらに図3に示したRemoveNode解析、を実行する（ステップ512）。この後、リストLからModify(n1→nx)とRemoveNode(nx)とを削除する（図5、ステップ513）。

#### 【0029】

ステップ510においてRemoveNode(nx)がリストL中に存在しない場合、単にノードn1の内容がnxに変更されたと解釈できるので、そのまま処理を終了する。差分計算部20は、以上の処理をリストL中の全てのModifyに対して行う。

以上のようにして、ツリーT、T'の差分が計算される。得られた差分データは、メインメモリ103等の記憶手段に一時的に保持され、XPath更新部30にて利用される。上記の3種類の解析にて実現されるように、本実施の形態では、ツリーTからツリーT'への変形において、ノードの移動として解釈できる操作はできるだけ移動操作moveとして検出することにより、後のXPathの更新処理に利用できるようとする。

#### 【0030】

XPath更新部30は、差分計算部20によるツリーTとツリーT'との差分の計算結果と、変更前文書Pに対応するXPath（以下、XPath（P）と記す）とを入力し、これらに基づいて変更後文書P'に対応するXPath（以下、XPath（P'）と記す）を生成し出力する。

図7は、XPath更新部30の機能構成を説明する図である。

図7を参照すると、指示情報生成手段であるXPath更新部30は、XPathを解釈する機能（XPath解釈器31）と、変更前後の構造化文書におけるノードの対応に関する情報を保持する機能（ノード対応表32）と、XPathを生成する機能（XPath生成器33）とを備える。そして、メインメモリ103やハードディスク105等の記憶手段から処理対象である更新前文書P、更新後文書P'、更新前文書Pと更新後文書P'との間の差分D、XPath（P）を入力し、上記の各機能を用いてXPath（P'）を生成する。生成されたXPath（P'）は、ハードディスク105等の記憶手段に格納される。

#### 【0031】

以下、XPath更新部30によるXPathの更新処理を詳細に説明する。

図8は、更新前文書Pと更新後文書P'との間のノード対応関係を示す図である。

X P a t h (P)は、各パスの階層（ロケーションステップ：LocationStep） $L_s(i)$  ( $i = 0, 1, 2, \dots, n$ )によって構成される。更新前文書Pにおいて、ロケーションステップ $L_s(i)$ の指示対象となるノードの集合（ノードセット：NodeSet） $S(i)$ が、X P a t h解釈器3 1の処理によって計算される。同様に、更新後文書P'において、ロケーションステップ $L_s(i)$ の指示対象となるノードセット $S(i)'$ が計算される。

一方、更新前文書P、更新後文書P'、更新前文書Pと更新後文書P'との間の差分Dから、更新前文書Pと更新後文書P'との間のノード対応関係を表すノード対応表3 2が作成される。作成されたノード対応表3 2は、図1に示したコンピュータ装置におけるC P Uの101のレジスタやメインメモリ103等に保持される。ノード対応表3 2の例を図10に示す。図示のノード対応表3 2によれば、例えば、更新前文書PのノードN 0が更新後文書P'のノードN' 0に対応しており、更新前文書PのノードN 0は更新後文書P'に対応するノードがない（すなわち構造化文書の変更によって削除された）ことがわかる。

#### 【0032】

このノード対応表3 2と、更新前文書Pにおけるロケーションステップ $L_s(i)$ の指示対象であるノードセット $S(i)$ に基づいて、ノードセット $S(i)''$ が求まる。

図9は、ノードセット $S(i)$ とノードセット $S(i)''$ との対応関係を示す図である。

ノードセット $S(i)'$ とノードセット $S(i)''$ との違いは、ノードセット $S(i)'$ が単純にパスのパターンを更新後文書P'に適用させて求めたノードセットであるのに対し、ノードセット $S(i)''$ は、差分情報により変更を追跡して求めたノードセットであるということである。また、ノードセット $S(i)'$ 、 $S(i)''$ とともに変更後文書P'におけるノードセットである。

#### 【0033】

次に、X P a t h生成器3 3により、このノードセット $S(i)'$ 、 $S(i)''$ を比

較して、XPath (P) におけるロケーションステップ Ls (i) を更新する。更新処理の詳細は後述する。この手順を  $i = 0 \sim n$  まで繰り返すことによりロケーションステップ Ls (j)' ( $j = 0, 1, 2, \dots, m$ ) を得る。このロケーションステップ Ls (j)' は直接、更新された XPath (P') を表す。

図11は、XPath生成器33によるXPath (ロケーションステップ) の生成手順を示すフローチャートである。

図11を参照すると、XPath生成器33は、まず、ノードセット S(i)' 、 S(i)" を比較する (ステップ1101)。そして、ノードセット S(i)' とノードセット S(i)" とが等しいか、またはノードセット S(i)' がノードセット S(i)" に含まれるのであれば、ロケーションステップ Ls (i) は変更する必要はなく、そのままロケーションステップ Ls (j)' として出力する (ステップ1102、1103)。

一方、ノードセット S(i)" がノードセット S(i)' に含まれる場合、XPath生成器33は、次にロケーションステップ Ls (j-1)' で示されるノードから、ノードセット S(i)" に含まれる各ノードへのロケーションステップを生成する (ステップ1103、1104)。

以上のようにして、変更後文書 P' に対応したロケーションステップが生成され、XPath (P) がXPath (P') へ更新される。

#### 【0034】

ところで、XPathの表記方法によっては、ステップ1104で生成されたロケーションステップを、所定の一般化ルールに基づいて一般化することにより、簡明な表現に集約することができる。もし、用意された一般化ルールによってはロケーションステップ Ls (j)' が生成できない場合は、警告を知らせるウィンドウや修正を促すウィンドウを表示するなど、エラー処理を行うと共に、ステップ1104で生成されたロケーションステップをそのまま出力することができる。

なお、ステップ1104におけるロケーションステップの生成は、例えば、下記文献1に開示された既知の戦略によって行うことができる。また、上記のロケーションステップの集約は、例えば、下記文献2に開示された既知の戦略によっ

て行うことができる。

文献1 : 2001/11/08: A Visual Approach to Authoring XPath Expressions Accepted for Markup Languages: Theory and Practice, Vol.3, No.2 (to appear). This is a paper originally published in the Proceedings Extreme Markup Languages 2001, pp.1-15, Montreal, Canada (14-17, August 2001).

<http://ares.trl.ibm.com/freedom/doc/extml2001/abe0114.html>

文献2 : 2001/07/13: XSLT Stylesheet Generation by Example with WYSIWYG Editing Accepted for the presentation at International Symposium on Applications and the Internet (SAINT 2002)

<http://ares.trl.ibm.com/freedom//doc/saint2002/saint2002.html>

### 【0035】

次に、具体的なツリーの変形例に基づいてXPathの更新方法を説明する。

図12乃至図14は、更新前文書Pと更新後文書P'の例を示す図である。なお、図には当該構造化文書P、P'のツリー構造が記載されている。

図12を参照すると、変更前文書Pは、ルートノードaが3つの子ノードbを持ち、そのうちの一番左のノードbが2つの子ノードcを持ち、そのうちの右側のノードcが子ノードbを持つツリー構造となっている。これに対し、変更後文書P'は、ノードcの下に位置するノードbがノードaの子に移動した構造となっている。

### 【0036】

ここで、変更前文書Pに対して、ノードaの子の2つのノードbを指すXPath (P) 「/a/b」が存在するものとする。「/a/b」という表現は、ノードaの子にある全てのノードbを指す。変更前文書Pが変更後文書P'に変更された場合、この「/a/b」をそのまま用いると、変更後文書P'においてノードaの子にある4つのノードbを指すこととなる。しかし、変更によってノードaの子に移動したノードbは、変更前文書Pにおいても存在し、かつこの時点では「/a/b」によっては指されていないノードであったので、変更後文書P'においても「/a

「/b」によって指されるべきではない。

#### 【0037】

本実施の形態においてXPath更新部30は、差分計算部20により計算された差分Dに基づいて作成されたノード対応表32により、図13に示すように、変更前文書PにおいてXPath(P)に指されていた3つのノードbが変更後文書P'における4つのノードbのうち左から数えて1番目から3番目までのノードbであることがわかる。そこで、XPath生成器33により、これらのノードb(すなわち変更前文書Pにおいても同じ位置に存在したノードb)のみを指すように、XPath(P')を生成する。すなわち、「/a/b」という表現を「/a/b[position() >= 3]」という表現に更新する。

#### 【0038】

また、図14を参照すると、図12と同様の変更前文書Pに対し、変更後文書P'は、ノードaの子の3つのノードbのうち、1つが削除されている。この場合、変更後文書P'において「/a/b」という表現で指し示されるノードの集合は、変更前文書Pにおいてかかる表現で指し示されるノードの集合に含まれる。すなわち、余分なノードを指してはいない。したがって、XPathの表現は変更する必要はない。なお、本実施の形態を用いた実施態様によっては、変更後文書P'において、指示するノードbの1つが削除されたことをユーザに通知するようにも良い。

#### 【0039】

以上説明したように、本実施の形態は、構造化文書が変更された場合に、変更前後の構造化文書の差分を検出し、これに基づいて対応するXPathを自動的に更新することを実現する。しかしながら、実際には、構造化文書やXPathを用いたシステムの開発者等の意図通りにXPathが更新されるとは限らない。また、XPathが自動的に更新された後、さらに修正を施したいと考える場合もあり得る。そこで、対話的に処理を進めるXPath更新ツールとして実施することも可能である。

図15は、このようなXPath更新ツールを備えたアノテーションシステムの例を示す図である。

図15において、アノテーションサーバ1500は、本実施の形態による文書解析部10、差分計算部20及びXPath更新部30に相当する機能を備えるものとする。これらの機能は、アノテーションサーバ1500を構成するコンピュータシステムにおいて、プログラム制御されたCPUの機能として実現される。アノテーション開発者が操作するコンソール1510のディスプレイ装置には、処理対象である構造化文書（例えばHTML文書）1511とXPathを更新するための対話用ウィンドウ1512とが表示されている。

#### 【0040】

アノテーションサーバ1500は、自身の管理下でアノテーションを付した所定の構造化文書1511が変更されると、コンソール1510のディスプレイに変更前後の構造化文書1511を表示すると共に、対話用ウィンドウ1512を表示して、この構造化文書1511の変更に応じてXPathを更新するかどうかをアノテーション開発者に問い合わせる。そして、対話用ウィンドウ1512の「Yes」ボタンがクリックされたならば、アノテーションサーバ1500の文書解析部10、差分計算部20及びXPath更新部30に相当する機能により、XPathが自動的に更新される。また、「Delete」ボタンがクリックされたならば、当該XPathが削除され、構造化文書1511のアノテーションが消去されることとなる。さらに、構造化文書1511において単に削除されたり変更されたりしたエレメント（ノード）に対するXPathは、参照不可能となる。この場合、アノテーションを付けたエレメントが削除されたことをアノテーション開発者に通知し、当該アノテーションを他のエレメントに付するかどうかを訪ねるメッセージを出力させるようにしても良い。

#### 【0041】

なお、以上の説明においては、XPathを用いてXMLやHTMLなどの構造化文書の要素を指示する場合について説明したが、本実施の形態は、XPath以外の任意の手段を用いて構造化文書の要素を指示する場合にも応用することが可能である。すなわち、構造化文書が変更された場合、本実施の形態に示した差分計算部20に相当する機能にて当該構造化文書の変更前後の差分を取り、当該構造化文書の要素を指示する手段（指示情報等）に応じた変形を行うことによ

り、かかる構造化文書の変更に対応してその要素の指示内容を適切に更新するこ  
とが可能である。

#### 【0042】

##### 【発明の効果】

以上説明したように、本発明によれば、構造化文書中の所定の要素が指示され  
ている場合に、この構造化文書が変更されても所望の要素を適切に指示し続ける  
ことができる。

また、本発明によれば、`XPath`を用いて所定の要素が指示されている構造  
化文書が変更された場合に、この構造化文書の変更に応じて`XPath`を自動的  
に更新する手段を提供することができる。

##### 【図面の簡単な説明】

【図1】 本実施の形態による`XPath`更新方法を実現するのに好適なコ  
ンピュータ装置のハードウェア構成の例を模式的に示した図である。

【図2】 図1に示したコンピュータ装置にて実現される本実施の形態によ  
る`XPath`更新システムの構成を示す図である。

【図3】 本実施の形態に好適な差分計算アルゴリズムによる処理を説明す  
るフローチャートであり、`InsertNode`解析を説明する図である。

【図4】 本実施の形態に好適な差分計算アルゴリズムによる処理を説明す  
るフローチャートであり、`RemoveNode`解析を説明する図である。

【図5】 本実施の形態に好適な差分計算アルゴリズムによる処理を説明す  
るフローチャートであり、`Modify`解析を説明する図である。

【図6】 本実施の形態に好適な差分計算アルゴリズムによる処理を説明す  
るフローチャートであり、`Modify`解析を説明する図である。

【図7】 本実施の形態における`XPath`更新部の機能構成を説明する図  
である。

【図8】 本実施の形態における更新前文書Pと更新後文書P'との間のノ  
ード対応関係を示す図である。

【図9】 本実施の形態におけるノードセットS(i)とノードセットS(i)"  
との対応関係を示す図である。

【図10】 本実施の形態にて用いられるノード対応表の例を示す図である

【図11】 本実施の形態のX Path生成器によるX Pathの生成手順を示すフローチャートである。

【図12】 更新前文書Pと更新後文書P'の例を示す図である。

【図13】 更新後文書P'に対応させてX Pathを更新した様子を示す図である。

【図14】 更新前文書Pと更新後文書P'の他の例を示す図である。

【図15】 本実施の形態を用いたX Path更新ツールを備えたアノテーションシステムの例を示す図である。

【図16】 一般的なツリーの差分計算の例を示す図である。

【図17】 一般的なツリーの差分計算の他の例を示す図である。

【図18】 XML文書の例を示す図である。

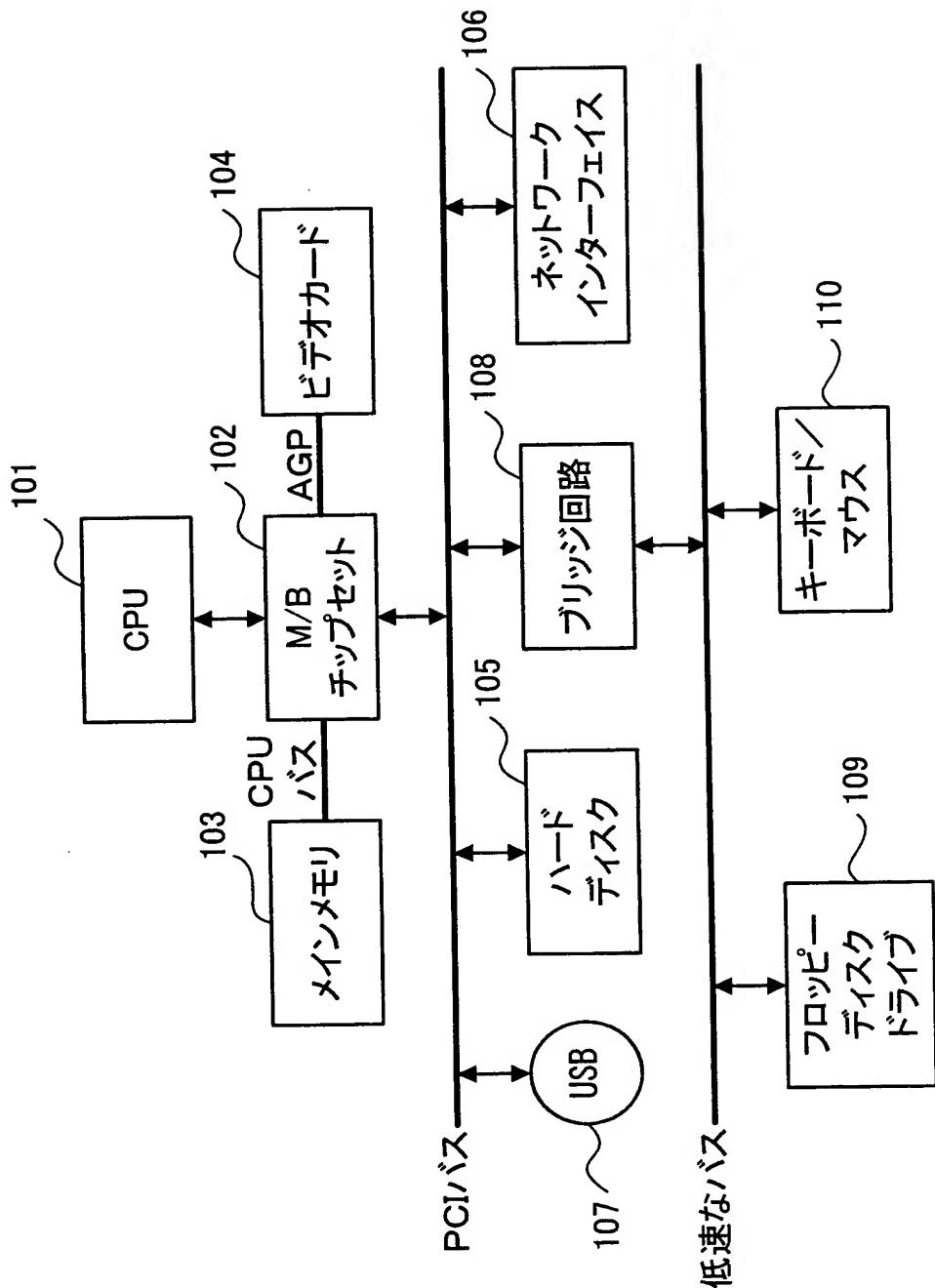
#### 【符号の説明】

10…文書解析部、20…差分計算部、30…X Path更新部、31…X Path解釈器、32…ノード対応表、33…X Path生成器、101…CPU、102…M/Bチップセット、103…メインメモリ、105…ハードディスク

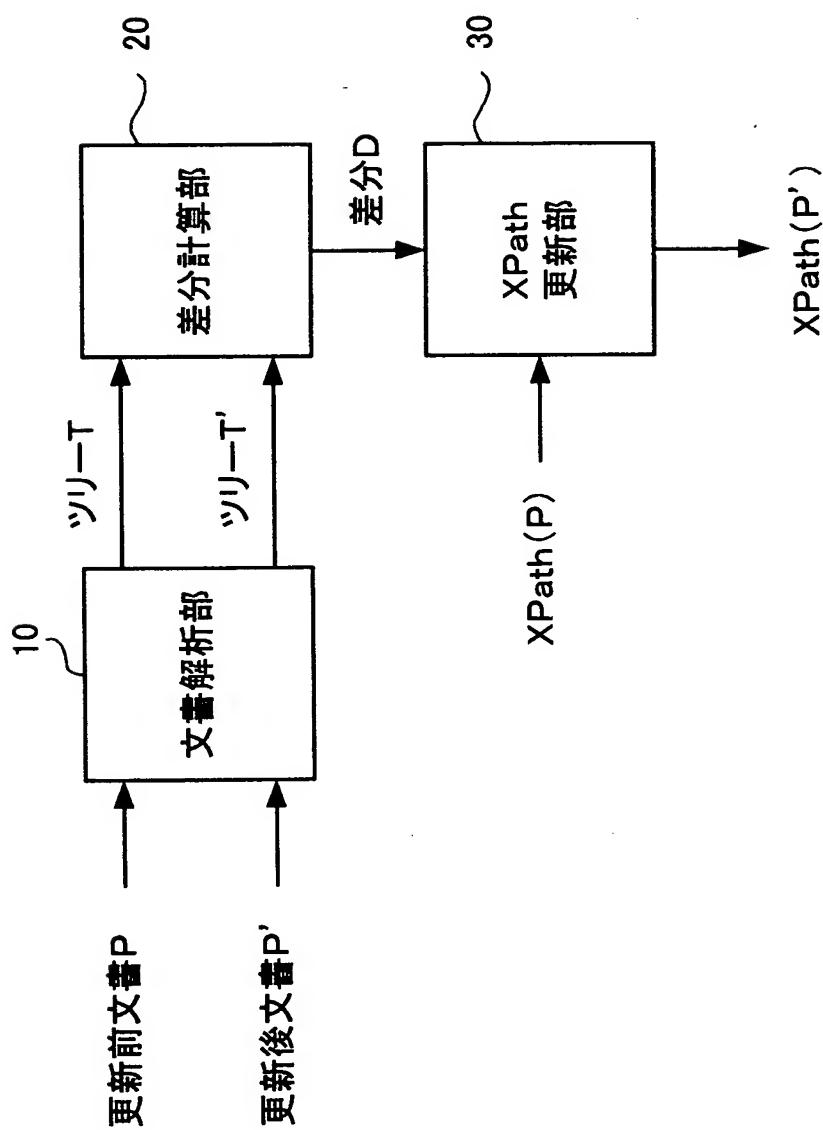
【書類名】

図面

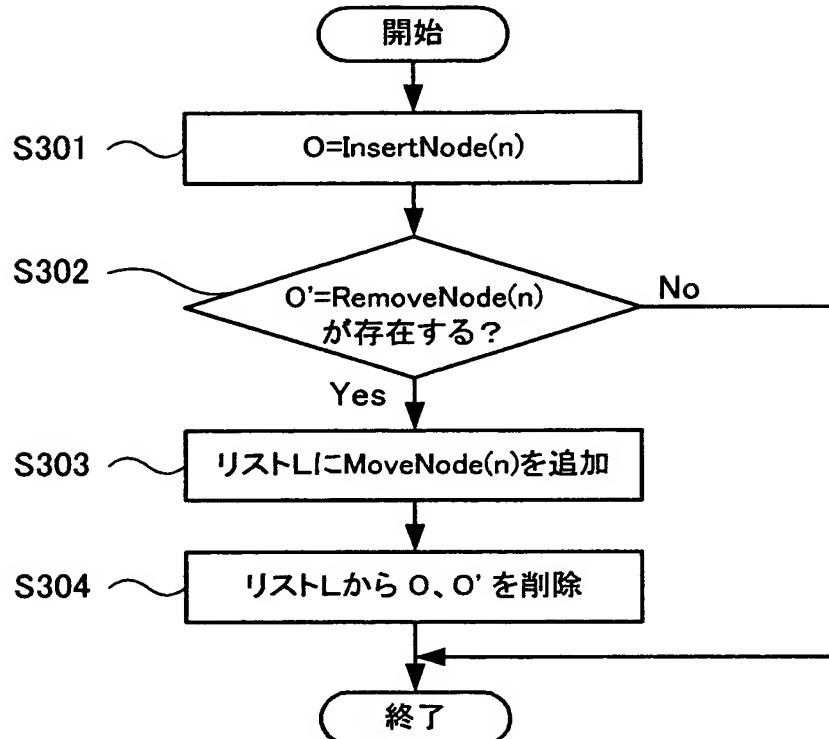
【図1】



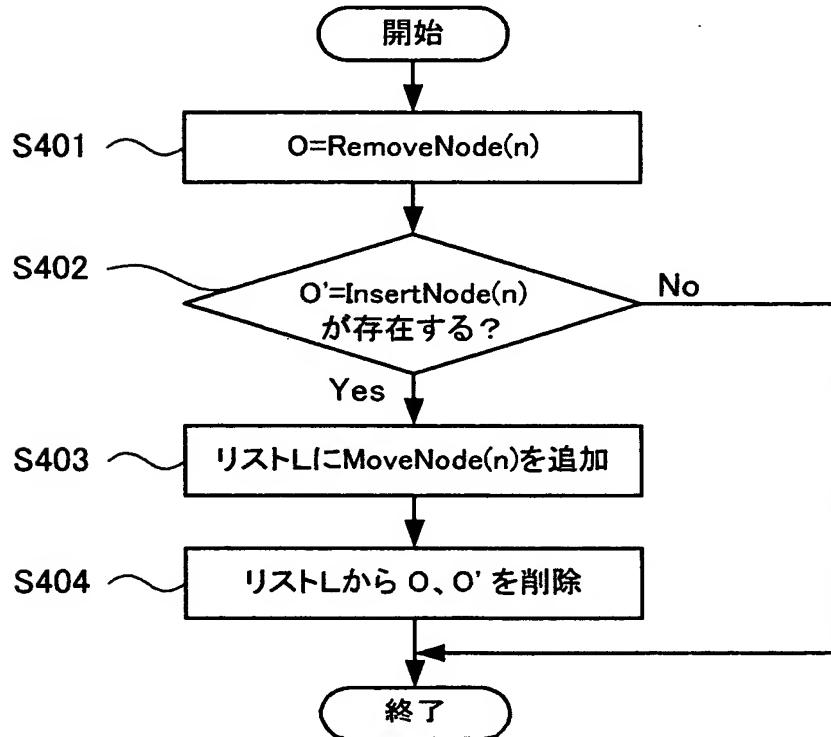
【図2】



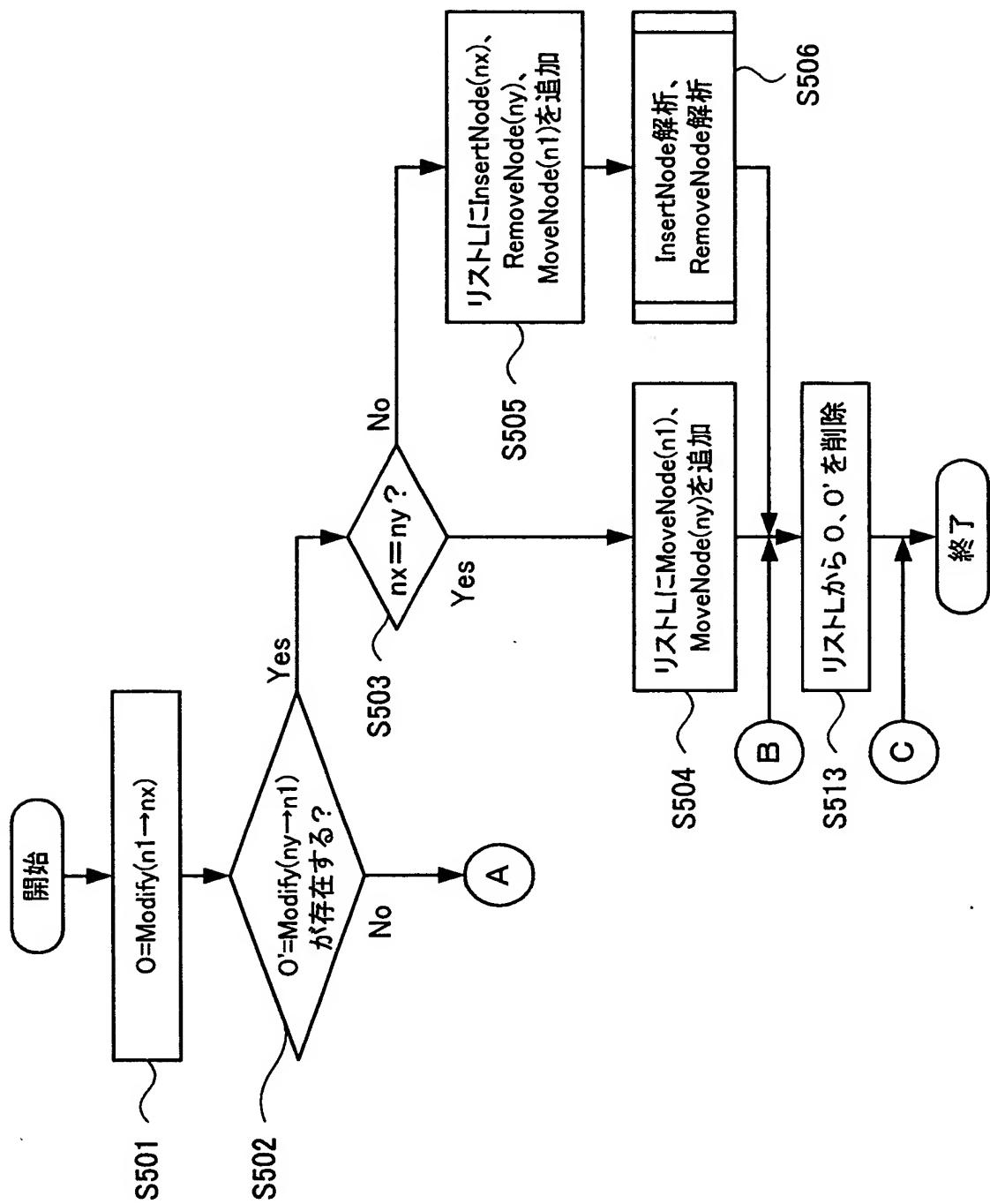
【図3】



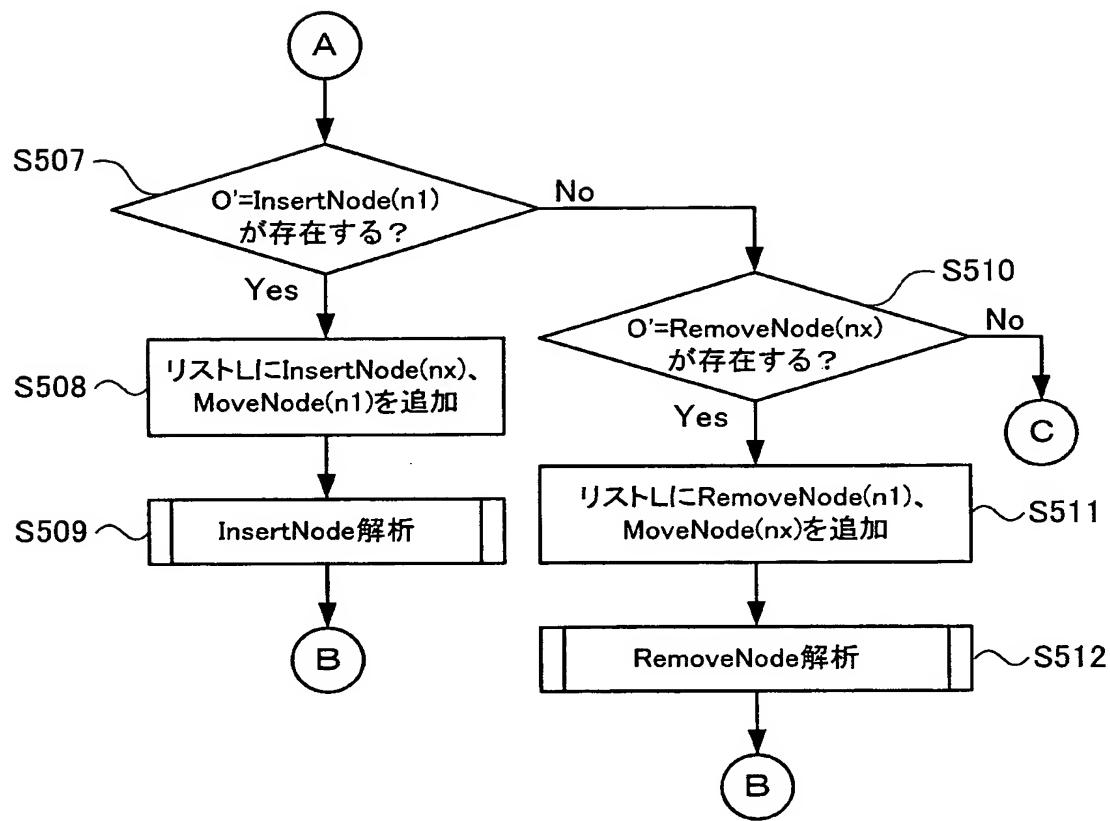
【図4】



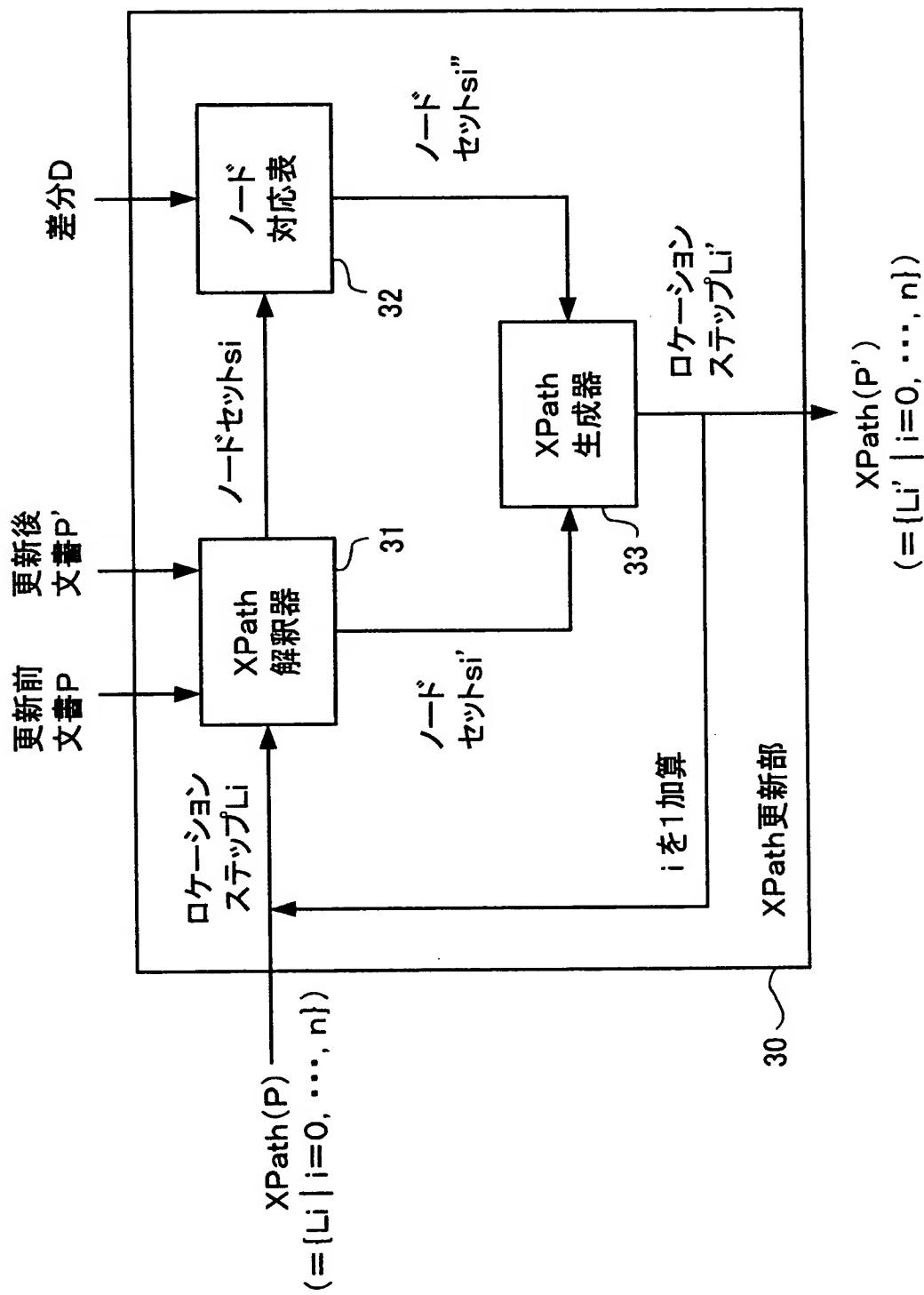
【図5】



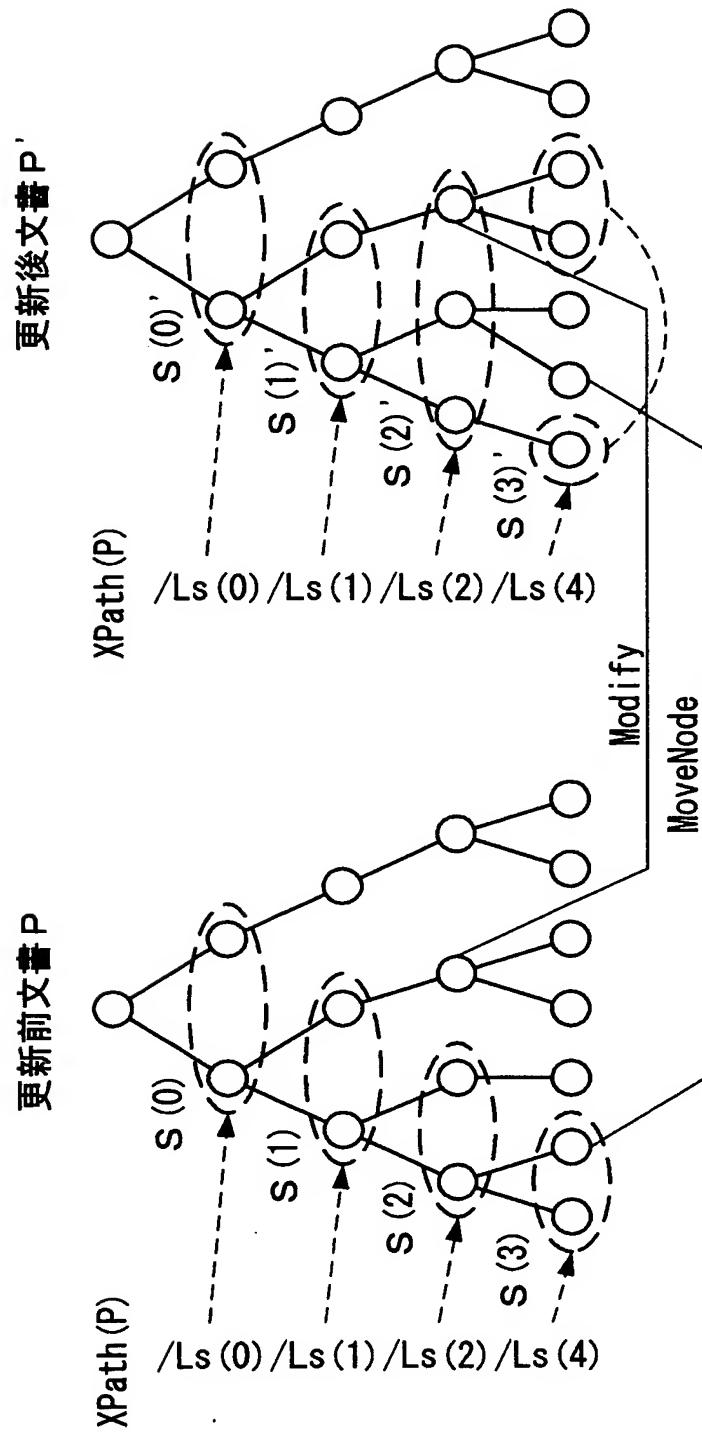
【図6】



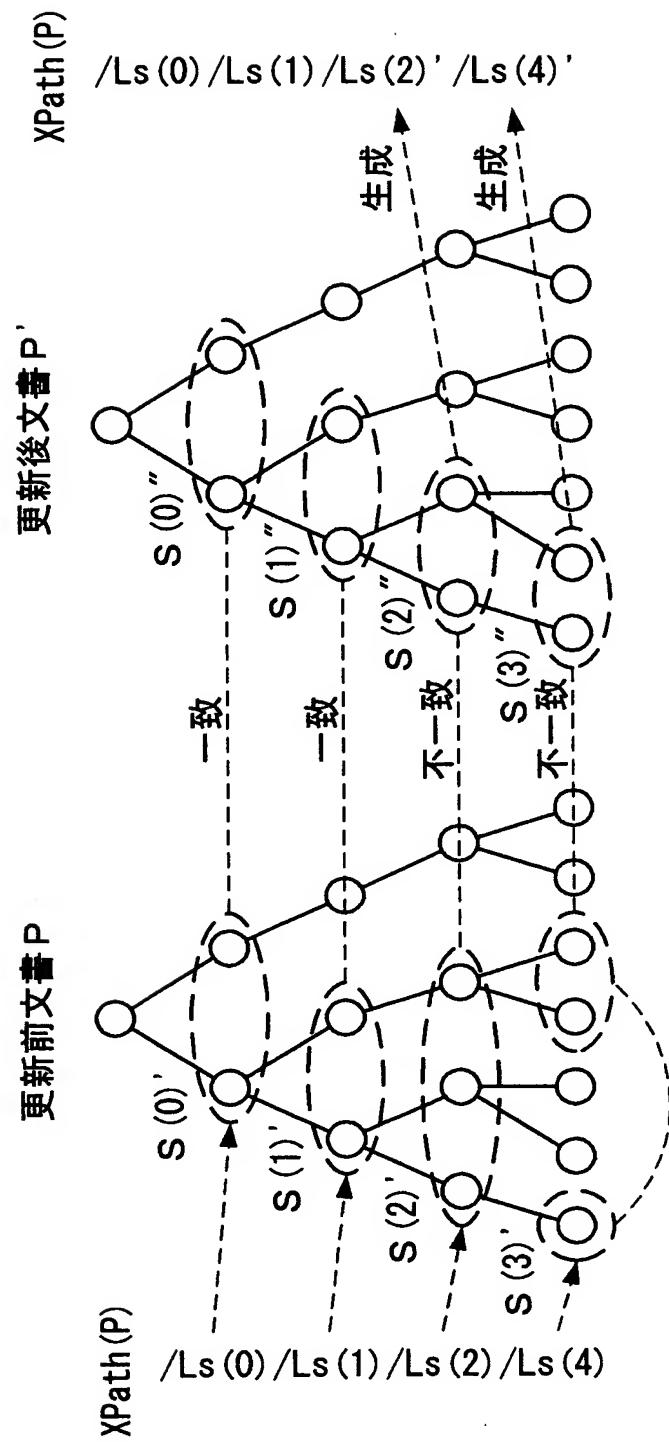
【図7】



【図8】



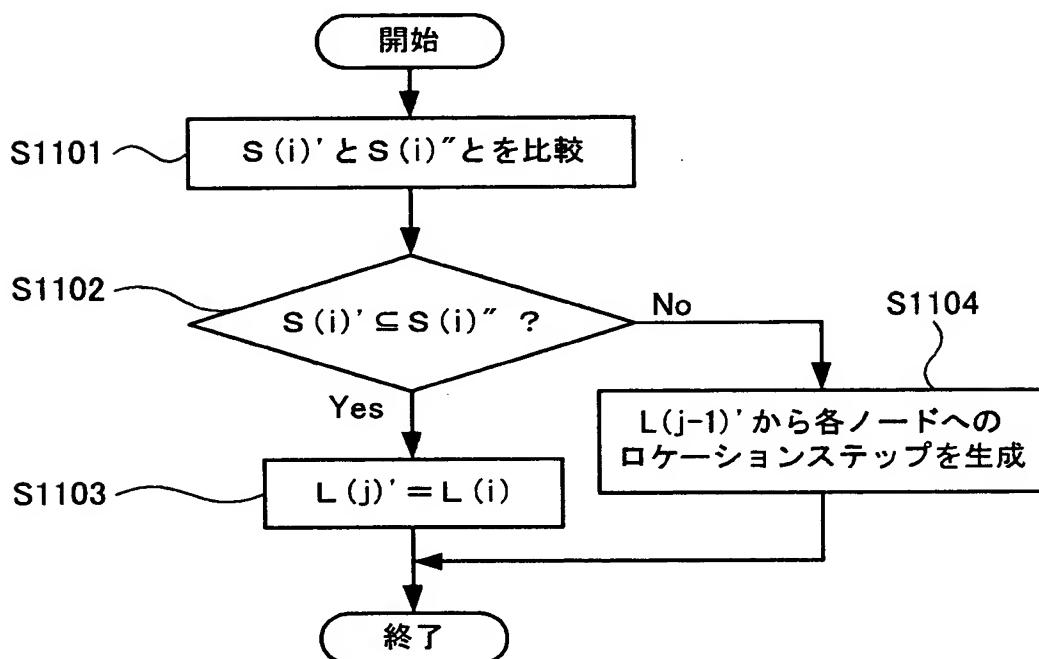
【図9】



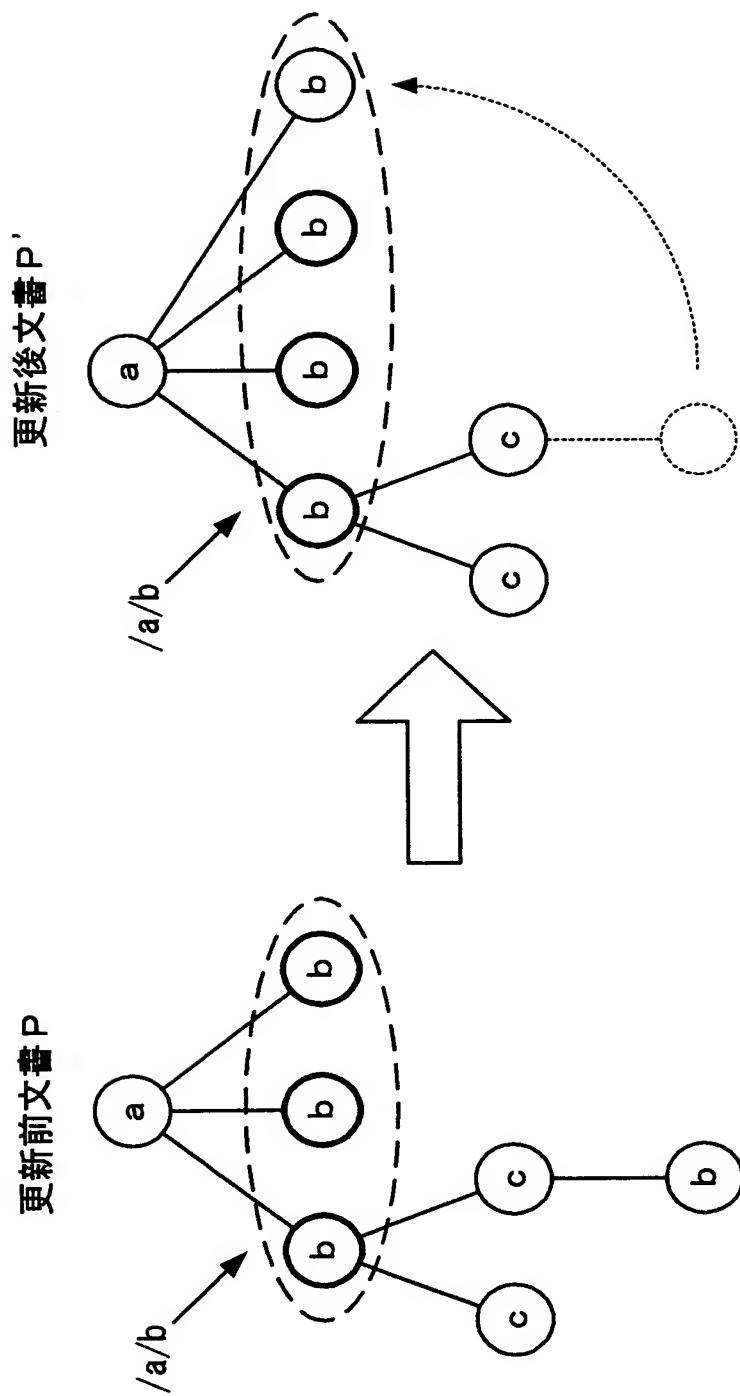
【図10】

更新前文書のノード	更新後文書のノード
N 0	N' 0
N 1	N' 1
N 2	N' 2
N 3	(non existent)
:	:
:	:
N n	N' n

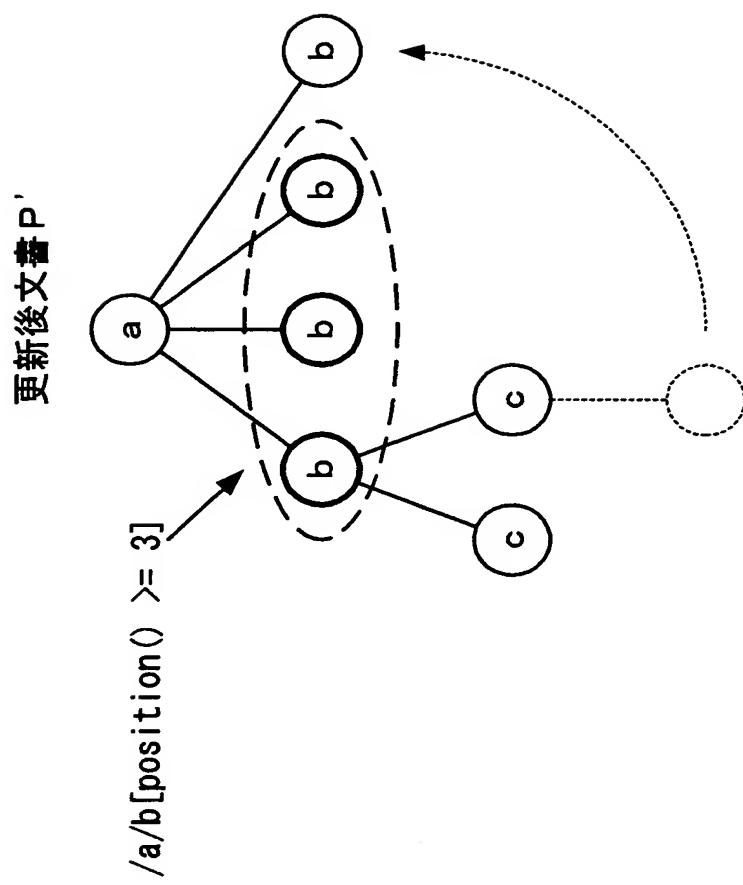
【図11】



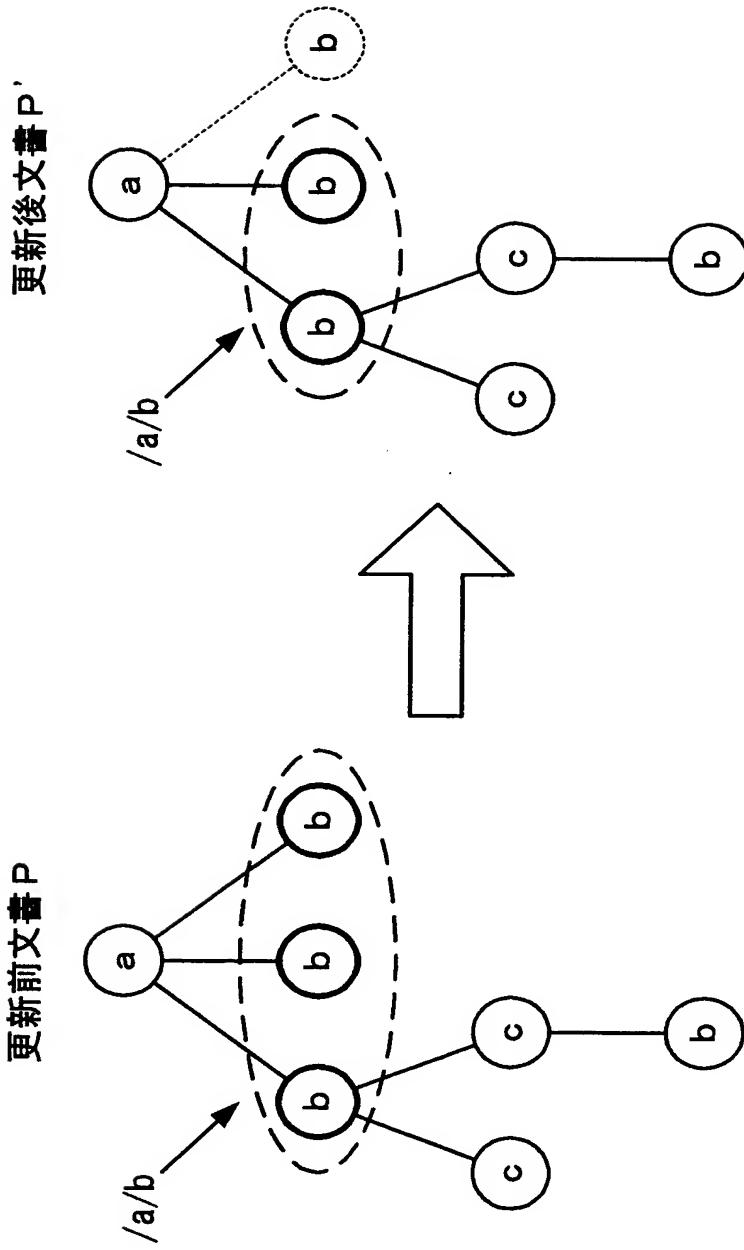
【図12】



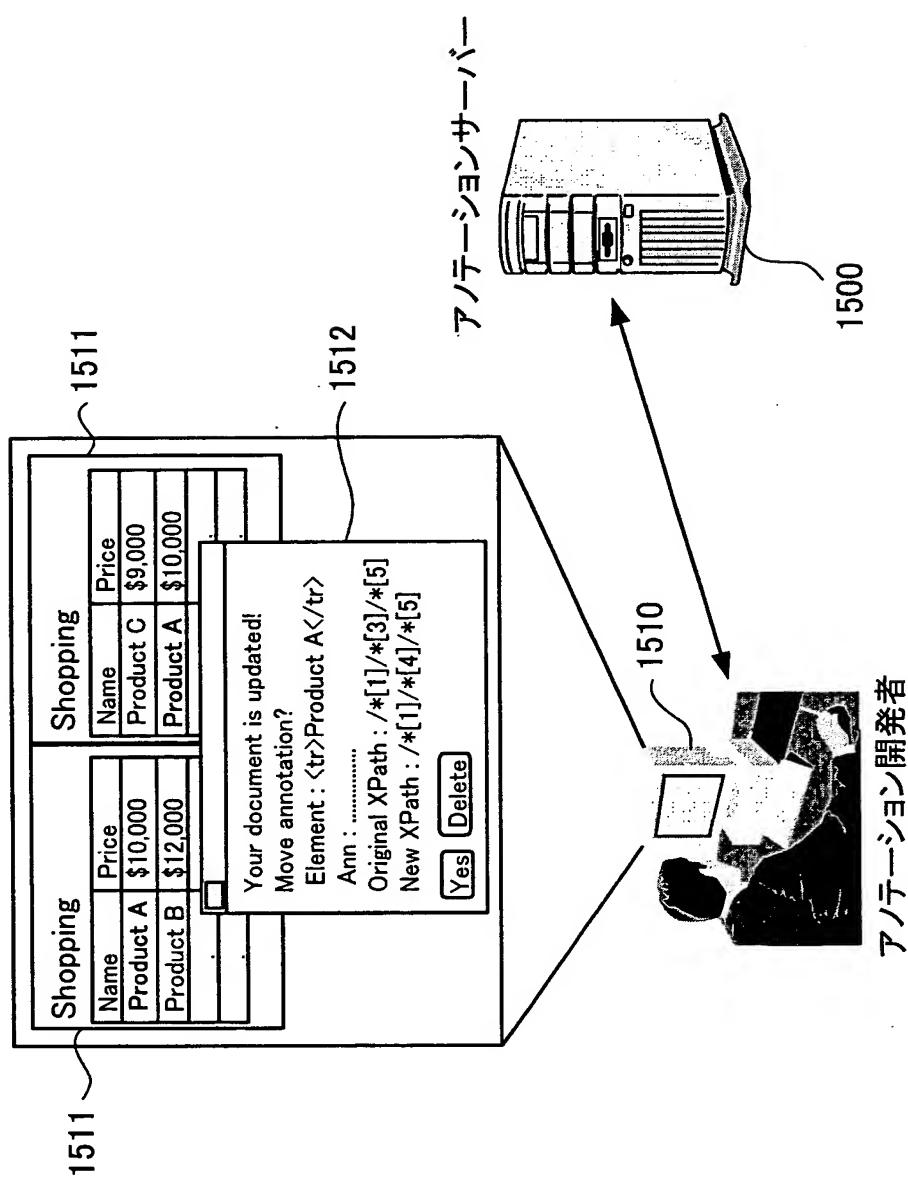
【図13】



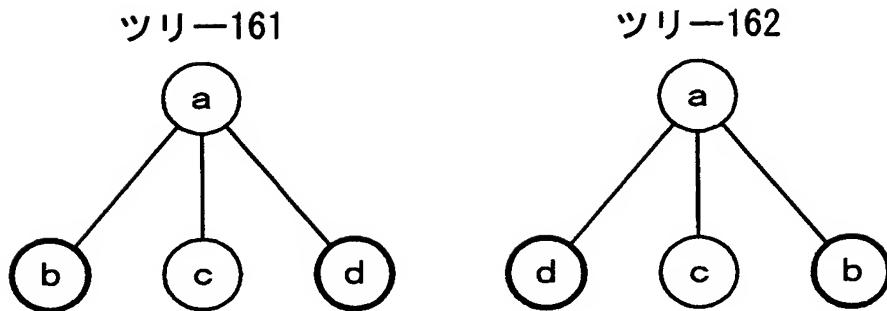
【図14】



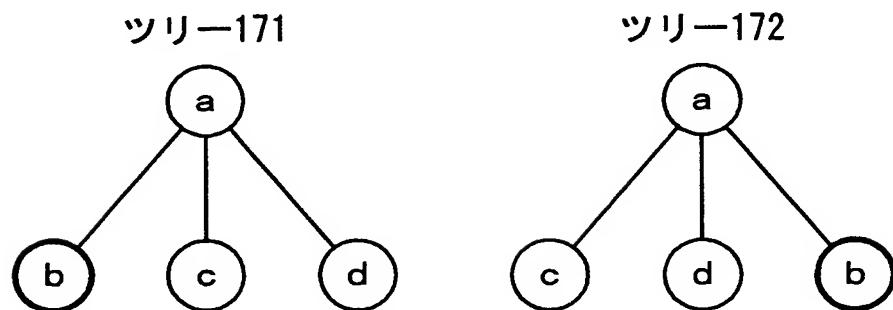
【図15】



【図16】



【図17】



【図18】

```
<a>
  <b>
    <d></d>
    Hello
  </b>
</a>
```

【書類名】 要約書

【要約】

【課題】 構造化文書中の所定の要素が指示されている場合に、この構造化文書が変更されても所望の要素を適切に指示し続けることができるようとする。

【解決手段】 複数の構造化文書の差分を計算する差分計算部20と、この差分計算部20により計算された差分に関する情報に基づいて所定の構造化文書における所定の部分を指示する指示情報から他の構造化文書中の対応する部分を指示する指示情報を生成するX P a t h 更新部30とを備える。

【選択図】 図2

## 認定・付加情報

特許出願の番号	特願2002-206202
受付番号	50201036733
書類名	特許願
担当官	小野寺 光子 1721
作成日	平成14年10月 3日

## &lt;認定情報・付加情報&gt;

## 【特許出願人】

【識別番号】	390009531
【住所又は居所】	アメリカ合衆国 10504、ニューヨーク州 アーモンク ニュー オーチャード ロード
【氏名又は名称】	インターナショナル・ビジネス・マシンズ・コーポレーション

## 【代理人】

【識別番号】	100086243
【住所又は居所】	神奈川県大和市下鶴間 1623 番地 14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	坂口 博

## 【代理人】

【識別番号】	100091568
【住所又は居所】	神奈川県大和市下鶴間 1623 番地 14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	市位 嘉宏

## 【代理人】

【識別番号】	100108501
【住所又は居所】	神奈川県大和市下鶴間 1623 番 14 日本アイ・ビー・エム株式会社 知的所有権
【氏名又は名称】	上野 剛史

## 【復代理人】

【識別番号】	100104880
【住所又は居所】	東京都港区赤坂 5-4-11 山口建設第2ビル 6F セリオ国際特許事務所
【氏名又は名称】	古部 次郎

次頁無

出願人履歴情報

識別番号 [390009531]

1. 変更年月日 2002年 6月 3日

[変更理由] 住所変更

住 所 アメリカ合衆国 10504、ニューヨーク州 アーモンク ニュー オーチャード ロード

氏 名 インターナショナル・ビジネス・マシーンズ・コーポレーション